

<https://helda.helsinki.fi>

Aatehistoria ja digitaalisten aineistojen mahdollisuudet

Tolonen, Mikko Sakari

2015-08-12

Tolonen , M S & Lahti , L M 2015 , ' Aatehistoria ja digitaalisten aineistojen mahdollisuudet ' ,
Ennen ja nyt : historian tietosanomat , Vuosikerta. 2015 , Nro 2 . <
<http://www.ennenjanyt.net/2015/08/aatehistoria-ja-digitaalisten-aineistojen-mahdollisuudet/> >

<http://hdl.handle.net/10138/232298>

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Ennen ja nyt - Historian tietosanomat

Aatehistoria ja digitaalisten aineistojen mahdollisuudet

<div>Kirjoittanut Leo Lahli</div> <div>Julkaistu 2015-08-11 11:08:11</div>
--

Aatehistoria on yleensä kiinnostunut laadullisista kysymyksistä.^[1] Alan tutkijat ohittivat esimerkiksi miten ihmisen moraalisuus tai yhteiskuntaa koossa pitävät mekanismit ymmärrettiin menneisyydessä. Tässä kontekstissa digitaalisten aineistojen merkitys on nähty ensisijaisesti aineistojen parantuneen saatavuuden kautta. Viime aikoina avoimen tiedon ”open access”-periaate on kuitenkin alettu nähdä laajempana kokonaisuutena, joka tarjoaa runsaasti uusia välineitä, lähestymistapoja ja mahdollisuuksia aatehistorian tutkimukseen. Yksinkertaistaneen voisi puhua vain avoimesta tiedestä. Kokonainen osaaminen aatehistorian ja digitaalisten ihmistieteiden alalla on jo kansainvälisesti vertailtuna korkeatasoista. Vahvistamalla uusia hyväksi koettuja tutkimuskäytäntöjä voidaan löytää tapoja nostaa entiseistään tekemisen tasoa. Erityisesti vahvan laadullisen tutkimustradition ja modernien tieteellisen laskennan yhdistäminen tarjoaa lupaavia ja ajankohtaisia mahdollisuuksia uusiin, kansainvälisesti merkittäviin avauksiin aatehistoriassa ja yleisemminkin ihmistieteiden alalla. Merkittävien tieteellisten tulosten saavuttaminen vaatii kuitenkin tieliasta, ennakko-luulotonta ja pitkäjänteistä tiedeenvälistä yhteistyötä yli perinteisten laitosrajojen. Avoimet verkostot ovat yhteyshenkilöiltä tarjonneet mahdollisuuksia digitaalisten aineistojen hyödyntämiseen tavalla josta voivat hyötymet monet eri osapuolet.

DIGIT AL HUMANIT IES – DIGITAALISET IHMIT IET EET

Digital humanities (DH) on maailmalla vakiintunut käsite, jolla tarkoitetaan erityisesti modernien tietojenkäsittelymenetelmien käyttöä humanistis-yhteiskunnallisessa tutkimuksessa koskien laajoja historiallisia, kirjallisia tai esimerkiksi kieltä ja kuvaa yhdistäviä digitaalisia aineistoja.^[2] Käsite on juurtunut tieteelliseen kielienkäyttöön 2000-luvun aikana. Tällä nimikkeellä ope roivat tieteelliset t edet ja professorit ovat pääasiassa vuosien 2010-luvulla perustettuja. Vastaavan suomenkielisen termin puuteessa Suomen Akatemia on päättänyt ke vällä 2015 viittaamaan ”digitaalisiin ihmistieteisiin”, joka kattaa sekä humanistisen alan että yhteiskuntatieteet.^[3] Termi on kuvaava ja lähellä englanninkielistä vastinettaan.

Kevään 2015 tehtien tulvissa digitalisaatiota, big dataa ja Suomen tulevaisuutta, kehitystä on seurattu rauhallisemmin historian alalla ja humanistisessa tutkimustraditiiossa. Alan kolmatisessa tutkimuksessa on kuitenkin tapahtumassa merkittävä käänne ja humanisteihin kohdistuu digitalisoidumispaineita niin alan sisällä kuin ulkopuoleltakin. Tietotekniikan käyttöä ei kuitenkaan ole humanistisessa tutkimustraditiiossa itsenäisä, eikä esimerkiksi perinteisten aineistojen digitoiminen sinänsä katso ta vielä kuuluvan alan varsinaiseen tutkimuksen piiriin. Jaottelu on haastavaa, sillä toisaalta tutkimusaineistojen keruu ja järjestäminenkin on kiistämättä keskeinen osa tieteellistä tutkimusprosessia.

Digitaalisia ihmistieteitä tehdään usein tekstiaineistojen pohjalta, mutta tutkimusaineistoksi käy siis kaikki kulttuuriperintö kuten esineet, kuva, ääni tai liike.^[4] Aineistojen kuvailutehtojen avulla on mahdollista tutkia esimerkiksi kirjoja materiaalisina objekteina tai kirjapainotoiminnan kehittymistä ja leviämistä historiallisena ilmiönä. Tutkimusaineistot voidaan jaotella digitaaliseen (esim. sosiaalinen media) tai muuten syntyneisiin (esim. digitoidut kirjat) aineistoihin. Keskityessä tässä kirjoituksessa ensisijaisesti ei-digitaalisiin syntyneiden aineistojen käyttöön joita aatehistorian tutkimustraditio erityisesti koskee. Digitaaliset ihmistieteet käsittelee piiriin katsotaan kuuluvan myös digitaalisuuden tutkimus ilmiönä. Voidaan tutkia miten kaupunkien toimintaan liittyvää avointa ja suljettua lähes reaaliaikaistakn dataa voidaan ymmärtää ja hyödyntää kaupunkien kehityksen kontekstissa. Kaupunkihin ja yhteiskuntaan liittyvää dataa on saatavilla kasvavassa määrin. Kyseessä on uudenlaisten tutkimusaineistojen ja -menetelmien käyttöönotto laadullisen tutkimuksen tueksi jo vakiintuneiden menetelmien rinnalle.^[5] Humanistien panos voi tuottaa tuoreita näkökulmia, kun tutkitaan esimerkiksi kaupunkien dynamiikkaan liittyvien ilmiöiden kehittymistä ja yhteisiä ihmisten käyttäytymiseen, hyvinvointiin, vuorovaikutukseen ja maailmankuvaan.

Digital humanities -käsite rinnastetaan toisinaan myös laskennalliseen yhteiskuntatieteeseen (Computational Social Science, CSS).^[6] Humanistis-yhteiskuntatieteelliseen tutkimukseen kenttä kattaaakin laajan kirjon erilaisia tutkimuskohhteita ja -menetelmiä. Tulee kuitenkin sisällyttää, että digitaaliset ihmistieteet -käsite viittaa tutkimukseen jossa uusia informaatiotieteiden menetelmiä hyödynnetään nimenomaan perinteisten humanistis-yhteiskuntatieteellisten kysymysten tutkimukseen muun muassa historiassa, lingvistiikassa ja muilla ihmistieteiden aloilla, kun taas laskennallinen yhteiskuntatiede on keskittynyt vahvemmin yhteiskunnan ja sosiaalisten verkostojen dynamiikkaa. Nuori, mutta jo paikkansa vakiinnuttanut laskennallisen yhteiskuntatieteen ala soveltaa ennakko-luultomasti esimerkiksi tilastollisen fysiikan menetelmiä sosiaalisten järjestelmien tutkimuksessa.^[7] Samalla sosiaalisuuteen liittyvät kysymykset ovat nousemassa entistä vahvemmin osaksi ”kovien” luonnontieteiden tutkimus kenttää. Tämä avaa näköaloja niin ihmistieteiden kuin informaatiotieteidenkin alalla ja uudistaa tutkimuskulttuuria. Samalla perinteiset tulokset so sialisuudesta saattavat kuitenkin suoraivastua tavalla joka on humanis- yheiskuntatieteen perinteelle vierasta. Kasvaessaan nuori luonnontieteistä poimittua laskennallinen yhteiskuntatiede elää ja monipuolistuu. Tähän liittyviä mahdollisuuksia tulee tunnistaa ja löytää yhteisiä kosketuspintoja muihin tieteenaloihin. Digitaalisen ihmistieteen tutkimus enasetelut ovat kuitenkin osin lähtökohdillaan erilaisia kuin vaikkapa laskennallisen yhteiskuntatieteen pyrkimykset ymmärtää yhteiskunnan toimintaa ja sosiaalista dynamiikkaa vaikkapa sovelletun fysiikan keinoin. Gadamer on esittänyt, että luonnontieteen menetelmiä ei voida sellaisenaan suoraan soveltaa ihmistieteisiin, koska näissä tulee ymmärtää inhimillisen toiminnan tutkimusidonnaisuus sekä jatkuva dialogi menneen, nykyisen ja tulevan välillä.^[8] Tähän sisältyy ripaus viisautta mistä kannattaa pitää kiinni.

Merkittävät tieteelliset läpimurrot vaativat syvällistä ymmärrystä uusien menetelmien avaamista mahdollisuuksista ja niihin liittyvistä rajoitteista. Ihmistieteiden tutkimuskysymysten on pohjaututtava perustutkimukseen ja tutkimustraditioon. Esimerkiksi kieliteknologia ei sellaisenaan joidenkin näkemysten mukaan lasketa digitaalseksi ihmistieteeksi, mutta toisaalta kieliteknologian sovelluksilla on merkittävää potentiaalia muun muassa aatehistoriallisen tutkimuksen keskeisenä tutkimusvälineenä ja alalta löytyy jo valmiiksi kotimaisia osaajia.^[9] Muita esimerkiksi ihmistieteissä jo hyödynnettäviä informaatiotekniikan menetelmiä ovat esimerkiksi sosiaalisen verkostojen analysointitekniikat ja muut datan kuvausmenetelmät kuten semanttiset verkostot. On selvää, että nousevan tutkimusalan alkuhuomassa tehtävistä koelluista vain osa kestää ja vakiintuu osaksi tutkimuskulttuuria. Näyttää kuitenkin varmalta, että modernien informaatiotekniikan menetelmien käyttö tulee uudistamaan ihmistieteiden tutkimusta sillä se tarjoaa uudenlaisia välineitä ja mahdollisuuksia systemaattiseen kokonaisuuksien muodoistamiseen. Laajojen tieteenalojen analyysi auttaa hahmottamaan yksittäisiä havaintoja laajemmista konteksteissa ja tällä tavoin saatavaa tietoa voidaan vertailla laadullisen tutkimuksen tarjoamaan kuvaan. Liikkuminen eri näkökulmien välillä ja erilaisten tutkimusperinteiden kohtaaminen voi synnyttää innovatiivisia tutkimusnäkökulmia ja tieteellisiä läpimurtoja. Samalla kun uusia menetelmiä käytetään klassisten kysymysten tutkimiseen, tulvat ne väistämättä muokkaamaan perustutkimuksen metodologista pohjaa ja avaavat uusia näkökulmia perustaan laatuisiin kysymyksiin.^[10]

”Digital humanities” -liikehdintä on lähivuosina saanut tuulta purjeisiin myös historian laitosten porteilla eri puolilla Suomea. Yhtäältä käsitteitä käytetään arkipäiväisesti kuvaamaan kaikkea digitaalisiin aineistoihin liittyvää tutkimusta ja toisaalta esimerkiksi Turussa, Porissa ja Tampereella on tehty tärkeitä aloitteita digitaalisen kulttuurin tutkimuksen alalla. Aiheen parissa toimivat saattavat kantaa humanistisen tutkimustradition voimaita taas kassaan jossa esiteltäen suuri paino yksittäisten sankaritarinajen teemoille laajemman yhteistyön sijasta. Näin tutkimuskentässä ei synny luontevasti kokemusta ja ymmärrystä monitieteellisydestä ja sen mahdollisuuksista, koska tutkimuskysymykset, aineistot ja menetelmät eivät ole kohdanneet tavalla jossa erilaisia taustoja omaavien tutkijoiden yhteistyö tuottaisi lisäarvoa. Kun yksittäinen tutkija pyrkii samanaikaisesti omaksumaan ja soveltamaan menetelmiä ja tietoa useita tieteenaloita, tekemisen määrä ja monimutkaisuus eivät voi korvata laatua. Usein sovrutaan liian helposti hätäisiin päätelmiin tai päädytään selvittämään yksittäisen tieteenalan näkökulmasta marginaalisia ongelmia kykenemättä edes kysymään – saatiik vastamaan – sellaisiin kysymyksiin, joilla olisi aloja pitkin lähtäimen vaikuttavutta tieteenalaa sisällä. Nykyiäne on laajalti otettava, että yksittäinen tutkija muuten kuin poikkeustapauksissa voisi hallita syvällisesti useiden tieteenalojen tutkimusperinnettä. Tarvitaa monitieteistä yhteistyötä, jonka tuottamaa lisäarvoa ei voida saavuttaa pelkästään jakamalla tehtäviä ja kasvattamalla te kijöiden lukumäärää.

Alan kehitykselle on keskeistä, että eri tieteenalojen tutkijat pystyvät ymmärtämään toistensa tarpeita ja vie tämään sujuvasti keskenään. Tämä vaatii kaikilta osapuolilta perusteellista yhteistyötahojen tutkimusperinteistä ja käsitteistä, sekä alalta mielenkiintoa oman tieteenalan ulkopuolisiin kysymyksiin ja eri tieteenalojen perinteistä ammentavan tutkimuksen tekoon. Ihmistieteiden tutkijat voivat osallistua laskennallisten menetelmien ideointiin ja soveltamiseen. Informaatiotieteilijät puolestaan voivat osallistua ihmistieteiden kysymysasettelujen muotoiluun ja rakentamiseen, mutta tutkimuksen laadun varmistamiseksi minä täytyy olla vankka asiantuntemus tutkimuksen kannalta keskeisillä aloilla. Sujuvan yhteistyön ja parhaiten tulosten saavuttamiseksi on ehdottoman tärkeää, että te kijoilla on jaettu näkemys tutkimushankkeen tavoitteista. Tätä tutaan parhaiten tekemällä tiivistä ja pitkäjänteistä yhteistyötä. Kokeuksia ja parhaita käytäntöjä voidaan ammentaa myös muilta tieteenaloilta, kuten laskennallisen biologian tutkimuksesta, joissa vastaavaa monitieteistä yhteistyötä on tehty jo pidempään.

AATEHISTORIAN JA KIELIT IET EEN LÄHEST YMIST APOJEN YHT ÄLÄISYYKSI ST Ä JA EROIST A

Kielen tutkija Franco Moretti on käyttänyt etäluennan (distant reading) termiä kuvataksaan maailmankirjallisuuden liikkeiden tutkimusta, jossa painopiste on yksittäisten dokumenttien sijasta laajempien trendien hahmottamisessa.^[11] Tämä kuvaa hyvin myös tavoitteita joihin pyritään monissa nykyisissä digitaalisten ihmistieteiden tutkimushankkeissa, kun etsitään laajoista tekstimassoista yhdistäviä te kijoita ja säännönmukaisuuksia vastauksena tiettyihin tutkimuskysymyksiin. Monimuotoisten digitaalisten aineistojen tulkinta vaatii selkeän asiantuntijapainoksen ja räätälöityjen laskennallisten työkalujen kehittämistä tutkimuksen tueksi.

”Big datan” käsitteelle ei ole vakiintunut suomenkielistä vastinetta, joten puhumme tässä tekstissä suurten aineistojen tiedonluhinnasta ja etäluennasta. Nykyään jo mantran lailla hoetaan, että dataintensivisten tieteenalojen näkökulmasta humanistisen tutkimuksen data on harvoin aistosti ”suuria” verrattuna esimerkiksi tähtitieteeseen, hiukkafysiikkaan tai molekyylibiologiaan. Tekstiorpukset joissa yhdistetään ja etäluetaan esimerkiksi kansainkirjallisuutta tutkijakäyttöön sekä muut vastaavat digitoidut aineistokokoelmat ovat usein kuitenkin niin laajoja ettei yksittäinen tutkija kykene mitenkään lukemaan tai tulkitsemaan koko aineistoa manuaalisesti. Samalla ihmistieteiden erityisluonne verrattuna luonnontieteen tutkimukohteisiin tuo omat lisähaasteensa aineistojen tilastolliseen analyysiin ja tulkintaan. Voilaissinkin Morettiä mukailien puhua esimerkiksi suurten aineistojen kvantitatiivisesta analyysistä. Korpus-tutkimuksessa on perinteisesti tähdytätty lähes täydellisten tutkimusaineistojen kokoomiseen. Siinä teksti on mahdollisimman tarkasti käsitelty ja rikastettu tutkimuksen vaatimaan muotoon. Etäluennassa taas aineiston ei tarvitse olla läheilläkään täydelliseltä, jotta siltä voisi jo olla merkittävää tutkimuskäyttöä.^[12] Sen sijaan pyrkimyksenä on, että aineisto sisältää riittävästi aineistia riittävästi tutkimuksen kannalta käytökehoista informaatiota joka voidaan kaivaa esiin sopivilla menetelmillä. Samalla kvantitatiivisen analyysin avulla aineistosta saadaan uutta tietoa esimerkiksi julkaisumääristä ja sanafrekvensseistä. Karkeakin aineisto voi näin olla tutkimuskieliseksi arvokasta pyrittäessä tunnistamaan laajoja tilastollisia säännönmukaisuuksia. Menetelmät voivat siinä huomattavakin lisäarvoa esimerkiksi epätarkkuuksia joitten vaikutus laajempien trendien havaitsemiseen on usein rajattu. Laajojen aineistojen kvantitatiiviseen analyysiin perustuva tutkimuksessa datasta voidaan tehdä myös odottamattomia havaintoja. Nämä voivat suunnata tutkimusta tavalla jota ei aluperin ole osattu ennakoida. Molemmilla lähestymistavoilla – huolella laadittujen pienempien aineistojen ja laajojen aineistojen yhdistäminen aineistojen tutkimuskieliseksi – on oma paikkansa Lähestymistavasta riippumatta aineiston siistiminen käyttökehoiseen muotoon on tutkimuksessa suhteellisesti yksi eniten aikaa vievistä vaiheista.

Aatehistoriassa laskennallisten menetelmien merkitys liittyy erityisesti niiden käyttöön tutkimuksen apuvälineenä jolla voidaan tunnistaa laajoista aineistoista mielenkiintoisia trendejä. Aineistojen analyysi usein liittyy laajempiin ja kielen käytön näkökulmasta ulkopuolisiin konteksteihin. Tämä poikkeaa sellaisesta lingvistästä tutkimuksesta jossa kielen muotojen tutkimuksessa automaattisoidulla menetelmillä tehdään jo nyt pitkäle vietyä johtopäätöstä liittyen esimerkiksi pronominien historialliseen käyttöön ja merkitykseen.^[13] Tämä esitys korostaa tarkoituksella tätä lingvistiikan ja historian tutkimuksen eroa. Kielitieteellä on automatisoitujen tutkimusmenetelmien käytössä selkeä etumatka. Tässä kritiikki kohdistuu erityisesti historian tutkimukseen itsensä ja kyyttömyyteen ottaa vastaavaa roolia teknologian kehityksessä mihin kielitieteissä on toisenlaisiin kysymysasetteluihin pystytty.

Aatehistoria on kehittynyt käsitteelliseen muutokseen josta tunnettu esimerkki on Quentin Skinnerin vaatimus keskittyä kirjoittajan retoriseen intentioniin, kun tämä tekee jonkun tietyn lingvistisen teon joka tulkitaan sen poliittisessa kontekstissa.^[14] Näin poliittisen kontekstin tunnistamisesta ja tulkitemisesta tulee keskeinen osa käsitteen käytön analyysiä. Skinner korostaa historiallisen kielen ymmärtämistä sen käyttö kontekstissa ja ettei käsitteitä välttämättä ole kytköksissä tiettyyn sanaan.^[15] Tämä erilainen ymmärrys kontekstista erottaa aatehistoriallisen tutkimuksen intressit useista muista sosio-lingvivistisistä tutkimusasetelmista. Edelluennan merkitystä ei kuitenkaan voida vähätellä edes skinneriläisen tutkimustradition jyrkimpinkään linjausten puitteissa. Skinneriläisen ajattelun ytimessä on jo vuosikaudet ollut erilaisten keskeisten poliittisten termien käytön laaja-alainen tunnistaminen ja tulkitus erilaisten poliittisten kontekstien kautta. Laajojen aineistojen käytössä korostuu mahdollisuus löytää uusin menetelmin historiallisia kiinnekohtia, jotka ovat aikaisemmin jääneet huomaamatta. Keskeinen rajoittava tekijä tässä on ollut pragmaattinen: yksittäisen tutkijan kyvyttömyys hallita suuria tekemänsöjä. Aikaisemmin laajat kontekstit on rakennettu useiden yksittäisten tutkijain tutkimusten varaan (kuten historian tutkimuksessa valitettavan usein muutenkin). Nyt tämän rinnalle on mahdollista saada kvantitatiivista tietoa esimerkiksi sanojen ja fraasien laajemmasta käytöstä.

Suhteessa Skinneriin, Reinhart Koselleckin esimerkin mukainen ja Suomessa vahvasti edustettu käsitehistoria on suuntaus jolla on eninomaista potentiaalia kasvaa ja kehittyä laajojen digitaalisten aineistojen käytön myötä.^[16] Käsitehistoria (*Begriffsgeschichte*) tutkii käsitteiden käyttöä, niiden muutosta sanastossa ja mikä sosiaalinen vaikuttavuus tiettyllä käsitteellä on ollut. Avainkäsite ei koostu ainoastaan sen suhteista muihin termiihin, vaan tulee ymmärtää kiistanalaisena ja poliittisten kamppailujen myötä sosiaalista maailmaa muokkaavana tekijänä. Näin sitä tutkitaan myös hieman eri lähtökohdista kuin skinneriläisiä intentioita. Tähän asti kuitenkin jopa diakronisen tutkimuksen tekeminen liittyen laajoihin historiallisiin aineistoihin on ollut vähintäänkin työlästä, mutta mallintamismahdollisuuksien kehityksessä käsitehistorian traditio pystymme hyödyntämään hyvän teoreettisen lähtökohdan. Automatisoidun edelluennan avulla voidaan paremmin kontrastoida tunnettujen ajattelijoiden kielenkäyttöä arkipäiväisempään julkiseen keskusteluun kuten sanomalehtiteksteihin. Näin voidaan paremmin hahmottaa sellaisia julkisen keskustelun konteksteja joissa klassikkotekstit ovat syntyneet. Voidaan olettaa, että tulevaisuudessa aatehistoriaa tehdäänkin kasvavassa määrin erilaisten sano- ja käsitteyrppäiden semanttisena analyysinä. Tätä kautta päästään tarkastelemaan uudestaan aatehistorian metodologista ydinkysymystä siitä miten "aate" tai "idea" voidaan määritellä ja tunnistaa. Syntyy myös tarve miettiä uudestaan myös kuopatuiksi luultuja ajattelijoita, kuten vaikkapa Lovejoita.^[17]

Uusista mahdollisuuksista huolimatta ei ole syytä vähätellä perinteistä lähiluentaan perustavaa aatehistorian tai filosofian historian tutkimusta, vaikka perinteisten metodologisten kysymysten rinnalle nousee uudenlaisia käytännön (sekä syvällisiä metodologisia) ongelmia, kun pystymme tutkimaan huomattavasti aiempaa laajempia aineistoja. Toisaalta laajemman kontekstin kytkeminen aineistojen analysointiin on haastavaa ja kohtaa väistämättä rajoituksia. Näin uudet informaatiotieteen menetelmät voivat tukea ja täydentää laadullista tutkimusta. On hedelmällöntä intää vastaan kun todetaan, että laajojen aineistojen analysointi tuo lähiluvun rinnalle tarkempaa tietoa julkisen keskustelun määrästä ja trendeistä, jotka ovat välttämättömiä uusien tulkintojen tekemiseksi.

KUVALUTIETOJEN MERKITYS JA AATTEIDEN MAANTIEDE

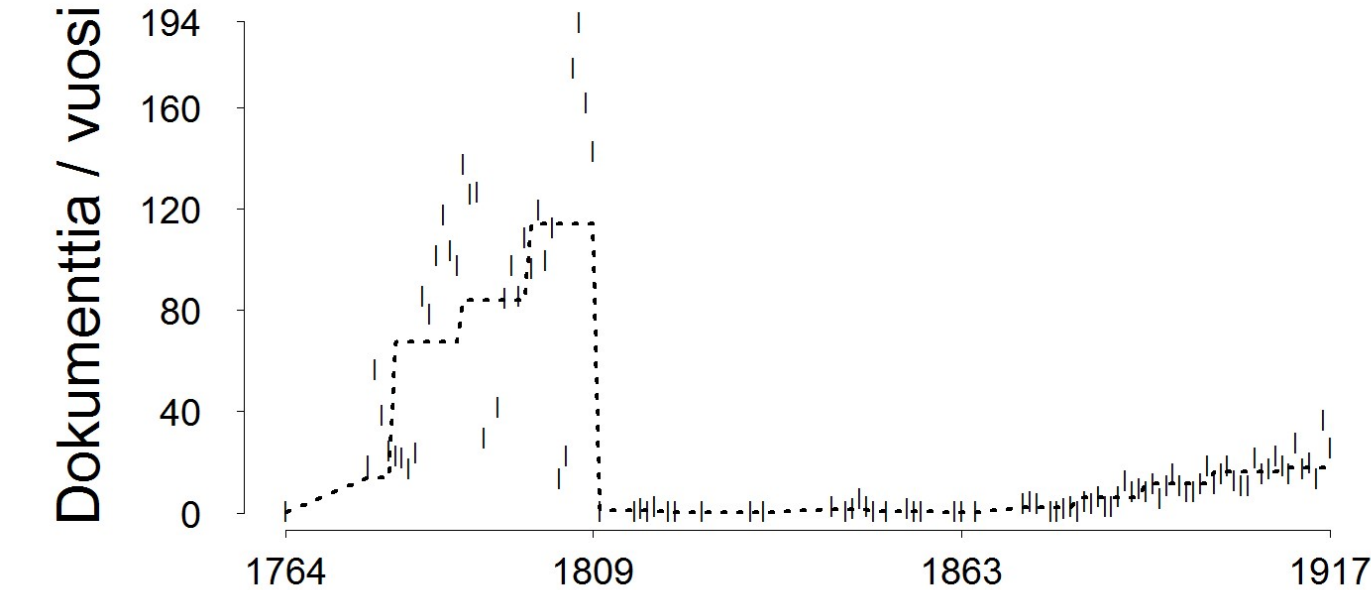
Laajojen tekstitietokantojen ohella aineistoja voidaan kartoittaa myös laajojen kuvailutietojen pohjalta. Näin voidaan tarkastella esimerkiksi kirjapainotoiminnan ja muun tiedontuotannon kehittymistä yhdistelemällä taustatietoa kirjojen painovuosista, -paikoista, -määristä, kirjoittajista ja aiheista. Lähteiksi käyvät monenlaiset aineistot: arkistosta rekisteriaineistoihin ja kirjastoluetteloihin. Yhtenäisen rakenteen omaavien kuvailutietojen automatisoitu jäsenely on usein helpompaa ja niitä voidaan tarkastella suoraviivaisemmillä tutkimusmenetelmillä kuin laajempia kokotekstitietokantoja, koska kuvailuaineistojen koko on rajatumpi ja rakennne selkeä. Tällainen lähestymistapa täydentää merkittävästi tekstipohjaisen tutkimuksen mahdollisuuksia ja komplementaarivien kuvailutietokantojen hyödyntäminen alan tutkimuksessa on saatettu tähän mennessä merkittävästi aliarvioitua.

Olemme tulkinne kuvailutietojen käyttöä muun muassa brittiläisen English Short-Title Catalogue (ESTC) -kirjastoluettelon pohjalta, jonka tiedot kattavat valloksen Britanniansa ja Pohjois-Amerikassa 1477-1800 painetuista kirjallisista dokumenteista. Tämän pohjalta pystymme tekemään suoraviivaista analyysiä aktuaalisen sosiaalisen muutoksen vaikutuksesta tiedontuotantoon (Kuva 1).

Kuva 1. Historia-aiheiset julkaisut Edinburghissa, 1540-1799. Kuvasta näkyy kuinka Edinburghissa julkaisuissa historia-aiheisissa dokumenteissa keskeisimmät yhteiskunnalliset muutokset vaikuttavat kirjastoluettelosta ilmeneviin julkaisumääriin vuosina 1540–1800. Näitä muutoksia, jotka näkyvät kaaviossa tummenneina, ovat: Englannin sisällissota (1642–1651), Stuart-suvun paluu valtaistuimelle (1660), Mainio vallankumous (1688–1689), Ison brittien kuningaskunnan syntyminen eli unioni (1706–1707). Huomionarvoista on myös ettei Yhdysvaltojen itsenäisyys (1776) aiheuta samanlaista julkaisuupjikkia.

Kotimaisen historian tutkimuksen kannalta kiintoisan esimerkin tarjoaa teollisten laskentakirjastojen laatuun suomalaisen kansallisbibliografian Fennican tutkimiseen, vaikka samalla tuleekin tiedostaa — ja korjata — Fennican nykyiset puutteet. Aineiston alustava analyysi osoittaa esimerkiksi suoraan historiallisiin tapahtumiin kytkytyvät tiedontuotannon huiput ja aallonpohjat Vaasassa, Helsingissä, Turussa ja Suomessa vuosina 1764-1917 (Kuvat 2-5).

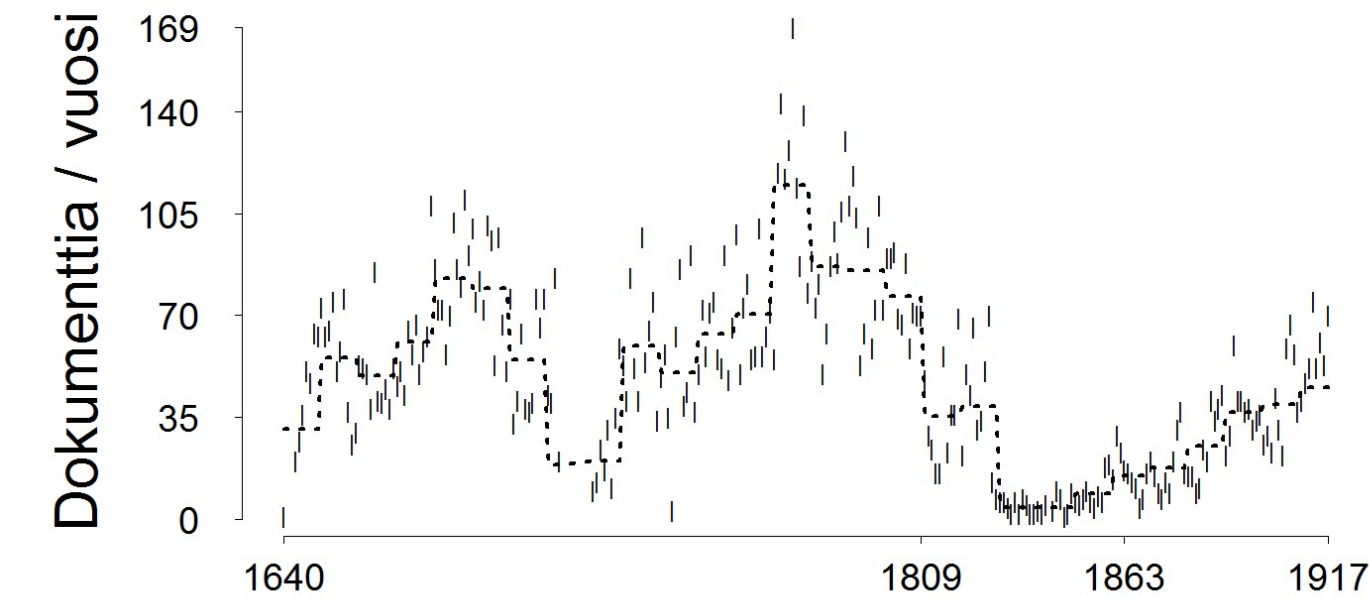
Julkaisutoiminta Vaasassa 1764-1917



Kuva 2. Julkaisutoiminta Vaasassa 1764-1917. Oheinen kuva näyttäisi viittaavan tiedontuotannon romahdukseen Vaasassa välittömästi Venäjän vallan alussa^[18]. Tämä saattaa osittain selittyä sillä, että painoimintaa harjoitti Vaasassa 1776-1804 yksi suurempi toimija (Georg Wilhelm Londicer) ja näyttäisi myös siltä, että hoivokeuden asiakirjojen julkaisuusperiaate ja painaminen muuttui Venäjän vallan myötä. Tämä kuvaa yksinkertaisella tavalla Suomen historiassa tapahtunutta julkisuuteen ja julkiseen keskusteluun liittyviä käännteitä. Suomen kansallisbibliografian Fennican perusteella julkisuuden kehitys painotuotannon muodossa näyttäisi olevan tilastollisesti näkökulmasta Venäjän vallan aikana hitaampaa kuin mitä yliopiston siirron ja Helsingin kehittämisen merkitystä korostanut historian tutkimus on aiemmin esittänyt.^[19] On kuitenkin alle viitattava, että mukana on muitakin selittäviä tekijöitä, kuten systemaattista keisarikunnan aikaisen painotuotannon liittyviä puutteita kirjastoluettelossa (on tunnettua, että Fennica ei ole täydellinen, vaan luetteloinnista puuttuu runsaasti erityisesti 1800-luvun ruotsinkielistä tuotantoa). On kuitenkin selvää, että aineistojen kvantitatiivinen analyysi tarjoaa tehokkaan tavan tunnistaa tiedontuotannon historiaan liittyviä trendejä. Lopulliset johtopäätökset tulee kuitenkin varmistaa myös muilla keinoin. Yksi

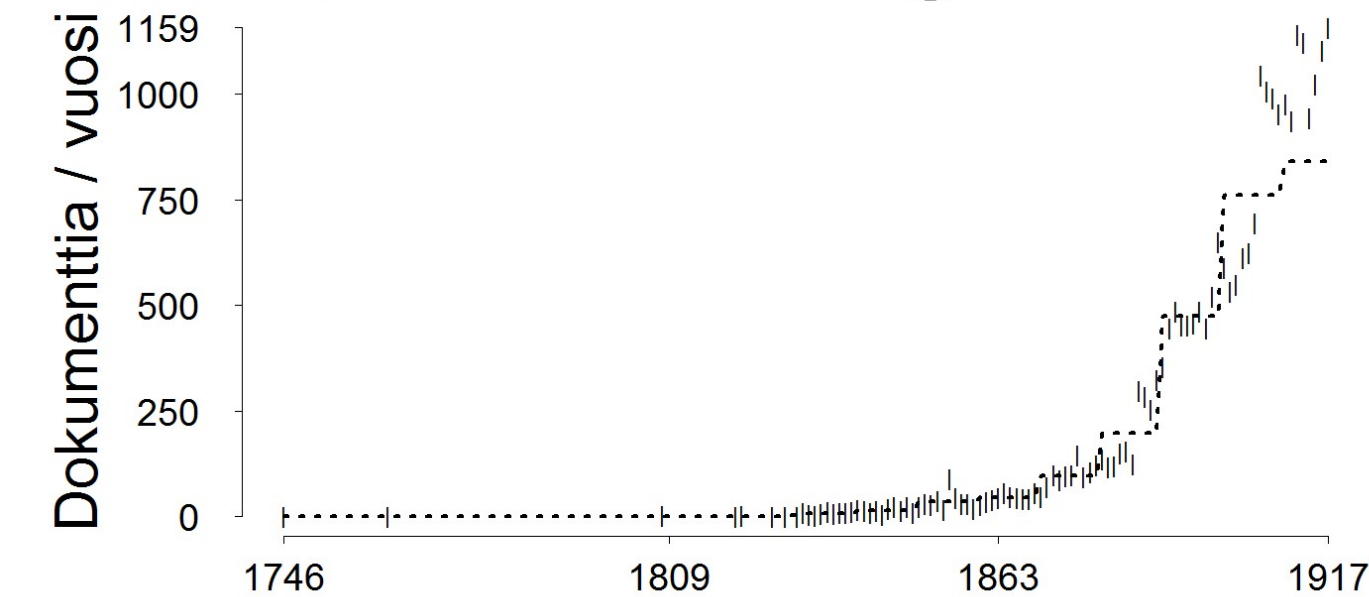
syy miksi tahdomme tässä julkaisussa tuoda nämä vielä vajavaiset tilastolliset näkökulmat esiin on, että emme yksinkertaisesti voi jättää odottelemaan kansallisbibliografian täydentymistä enää seuraaviksi vuosikymmeniksi (ks. esim. huomio Jyrki Hakapää, Kirjan tie lukijalle. Kirjakauppojen vakiintuminen Suomessa 1740-1960, SKS, 2008, s. 380), vaan tämä tärkeä luettelointityö olisi tutkimuksellista syistä kyettävä viedä loppuun mahdollisimman pian.

Julkaisutoiminta Turussa 1640-1917



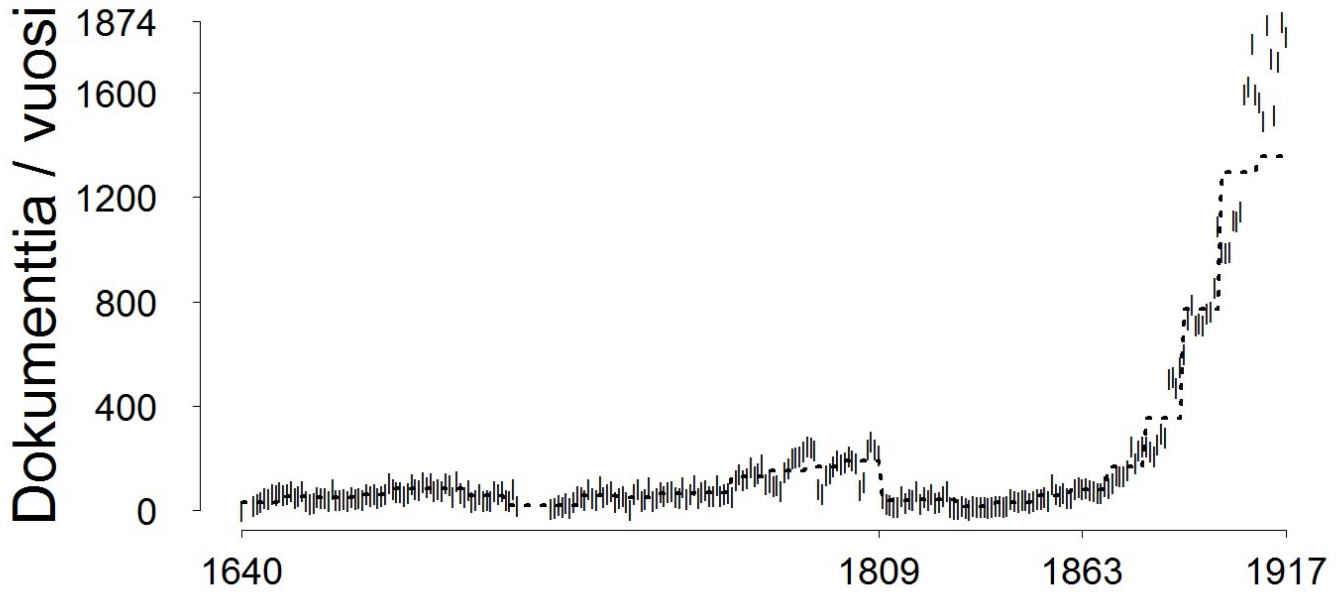
Kuva 3. Julkaisutoiminta Turussa 1640-1917. 1700-luvun kehitys näyttäisi seuraavan eurooppalaista kasvavaa trendiä, kirjapainotoiminnan kehitys 1820-luvulta lähtien on taas nykyisten luettelointitojien valiossa hidasta.

Julkaisutoiminta Helsingissä 1746-1917



Kuva 4. Julkaisutoiminta Helsingissä 1746-1917. Visuaalinen esitys Helsingissä tapahtuneesta kirjapainotoiminnan kehityksestä viittaa siihen, että jos Fennican luettelointiin on ollenkaan luottamista, niin ennakkosensuurin osittaisen poistumisen vaikutukset 1850-luvulla ovat yllättävän vähäiset ja varsinainen painotoiminnan kehitys lähtee käyntiin vasta huomattavasti myöhemmin. Yliopiston siirtäminen Helsinkiin näyttää olleen julkaisutoiminnan kannalta, jos ei lamauttava, niin ainakin hidastava toimenpide.

Julkaisutoiminta Suomessa 1640-1917



Kuva 5. Julkaisutoiminta Suomessa 1640-1917 kertoo siitä, että 1700-luvun lopun ruotsalainen kehitys näyttäisi suunnilleen samalta kuin yleinen eurooppalainen painotoiminnan nousu. Vuoden 1809 jälkeen toiminta Suomen alueella taantui ja kestäi aina 1890-luvulle asti, että lukumääräisesti painotoiminnassa nousuaan samalle tasolle kuin Ruotsin vallan aikana. Tämä kuva korjaantuu ainakin osittain kansallisbibliografian täydentyessä.

Kartoittamalla lisäksi vaikka ruotsalaisiin tai venäläisiin paikannimiin tehtävien viittausten määrää ja laatua suomalaisissa sanomalehdissä, voimme saada uutta tietoa suomalaisen julkisen keskustelun rakenteesta ja kehityksestä historian eri vaiheissa.^[20] Samaa voidaan tehdä paikallisessa mittakaavassa tutkimalla viittauksia naapurikuntien ja suurempien kaupunkien välillä.^[21] Kansallisbibliografian historiallisten henkilöiden nimien paikantaminen ja kartoitus voi myös osaltaan kehittää suomalaisen julkisen sfäärin tutkimusta. Näin uudet informaatiotekniikan menetelmät voivat antaa aatehistorian tutkimuksen käyttöön uudenlaista tutkimusmateriaalia ja tutkimuksellisia näkökulmia.

Alkuperäiset kuvailutiedot eivät sellaisenaan soveltu tilastolliseen analyysiin, vaan ne pitää ensin jalostaa sopivaan muotoon. Tämä sisältää mm. maantieteellisten ja henkilönimien erilaisten kirjoitusmuotojen yhdistämistä, sanomien kirjoittajien erottelua, kirjoitusvirheiden käsitteilyä, tai kirjojen sivumäärän arviointia bibliografisten standardimerkintöjen pohjalta. Tietoja voidaan myös täydentää ja rikastaa muiden lähdeaineistojen nojalla. Kirjoittajista voidaan esimerkiksi selvittää puuttuvat elinvuodet, sukupuoli, muuta taustatietoa, painopaikkojen perusteella painomaa ja niin edelleen. Valtiosa työstä voidaan ja tulee automatisoida, kun tutkittavien aineistojen määrä kasvavat kymmeniin tai satoihin tuhansiin dokumentteihin. Samalla analyysien tilastollinen voima kasvaa ja satunnaisten virheiden vaikutus johtopäätöksiin häviää. Kaikkea ei voida kuitenkaan automatisoida. Juuri tästä syystä joustavaan lähdekoodiin perustuvat räätälöidyt laskentakirjastot tarjoavat erinomaisia mahdollisuuksia alan tutkimukseen verrattuna valmiisiin ohjelmistokokonaisuuksiin, jotka usein olisivat etä aineisto on jo valmiiksi saatettu tilasto-ohjelmiston vaatimaan muotoon. Tämä on merkittävä puute, sillä käytännön kokeuksen perusteella aineistojen siittäminen ja laadunvarmistus ohjelmistojen ja tilastollisten analyysien vaatimaan muotoon on yksi työläimmistä ja eniten aikaa vievistä työvaiheista. Tieteellisen ohjelmoinnin hyödyntäminen auttaa automatisoimaan kaiken mikä automatisoitavissa on. Tieteellinen laskentakirjasto tarjoaa nimensä mukaisesti kokonaisen hyvin dokumentoituja menetelmiä tutkimuskokteeni erityispiirteet huomioivien tietojenkäsittelytyövälinein. Laskentakirjastot auttavat automatisoimaan merkittäviä osia aineistojen tilastollisesta analyysistä, mutta valmisohjelmistoista poiketen antavat samalla tutkijalle merkittävästi laajemmat vapaudet yhdistellä ja täydentää menetelmiä halutulla tavalla, tutkimuskokteeni erityispiirteet huomioiden. Tämä tarjoaa tutkijalle sujuvan pääsyn sekä aineistojen jalostamiseen että tilastollisen analyysin ja visualisointien kannalta keskeisiin työkaluihin. Laskentakirjastoja ylläpidetään usein avoimen lähdekoodin palveluissa, jolloin koko työvirta voidaan toteuttaa läpinäkyvällä ja toistettavalla tavalla, tieteellisen avoimuuden periaatteita kunnioittaen (ks. kuva 6).

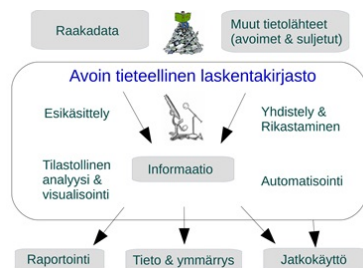
Tiedonlouhinnan vaihtoehtoja aatehistorian näkökulmasta

Laajojen aineistojen tekstilouhinta

- **Tavoite:** ymmärtää käsitteellistä muutosta ja kielen käyttöä
- **Aineistot:** tekstitietokannat (ECCO, EEBO, digitoidut sanomalehdet jne.)
- **Potentiaali:** kvantitatiivinen evidenssi laadullisen tutkimuksen tukena ja uudet havainnot; etäluenta
- **Käytäntö:** alkuperäisaineistot harvoin avoimesti saatavilla tai sidottu rajoitettuihin käyttöliittymiin
- **Skaalautuvuus:** rajattu; data on kooltaan massiivista ja voi vaatia huomattavaa jalostamista; aineistosta ja menetelmistä riippuvainen;
- **Metodologinen näkökulma:** aatehistorian sovelluksia toistaiseksi hyvin vähän

Kuvailutiedot tilastollisena työkaluna

- **Tavoite:** ymmärtää yleisiä trendejä historiallisten ilmiöiden dokumentoinnissa
- **Aineistot:** kirjastot ja arkistot tulvivat erilaisia kuvailutietokantoja (metadata)
- **Potentiaali:** tutkimuskäyttö aliarvioitu
- **Käytäntö:** saatavuudessa ongelmia, mutta vähemmän kuin tekstiaineistojen kohdalla
- **Skaalautuvuus:** erinomainen; data on rakenteista ja sen koko on rajattu
- **Metodologinen näkökulma:** tarjoaa selkeän lähtökohdan uusien käytäntöjen soveltamisessa humanistiseen tutkimukseen.



Kuva 6. Datasta tietoon.

Alkuperäisen tutkimusaineiston jalostaminen tiedoksi ja lopulliseksi julkaisuksi on monivaiheinen prosessi, jossa tarvitaan välineitä aineistojen järjestämiseen, yhdistelyyn ja tilastolliseen analysointiin. Merkittäviä osia tästä työstä voidaan automatisoida laatimalla tarkoituksen suunnattuja tieteellisiä laskentakirjastoja. Laskentakirjasto on hyvin dokumentoitu kokonaisu tutkimuskokteeni mukaan räätälöityä laskenta-algoritmeja, joilla voidaan automatisoida monia rutiinomaisia työvaiheita ja luoda tähän yhteisiä käytäntöjä. Samalla lähdekoodin avoin julkaisu verkossa tekee tutkimusprosessista toistettavan ja läpinäkyvän, mahdollistaa niin datan kuin välineiden tehokkaan jatkokäytön muissakin yhteyksissä, ja edistää laajempaa

tutkijaverkoston osallistumista kehitystyöhön.

Ideaalitilanteessa digitaalisten ihmistieteiden tutkimushankkeita toteuttavat tutkimusryhmät kootaan vastaamaan selkeisiin tutkimuskysymyksiin ja tutkimusmenetelmien kehityä ja soveltamiseen tapahtuva ryhmän jäsenen yhteistyönä. Tärkeää on työn suunnittelu niin, että muualla laajatuja parhaita tieteellisiä käytäntöjä voidaan siirtää tehokkaasti humanistis-yhteiskunnalliseen tutkimuksen piiriin. Hankkeeseen osallistuvien humanistien tulee omata riittävää ymmärrys alan vaatimuksista ja itseluottamusta niihin vastaamiseen. Tekstitaineistojen tutkimukseen liittyy kaksi toisaan täydentävää lähestymistapaa, joita ovat rakenteellaton tekstin louhinta sekä erilaisten aineistojen kuvallistietojen tutkimuskäyttö, joka pitää sisällään esimerkiksi kirjasto-, arkisto- kuin myös museoluettelointit (ks. ylläoleva kaavio). Näiden hyödyntäminen tutkimuksessa ei poista perinteisen lähtelun merkitystä, vaan täydentää sitä ammentamalla aiheita laajemmasta perspektiivistä.

Tekstilouhinnan potentiaali on valtava. Jos aatehistorioitsija on esimerkiksi kiinnostunut tutkimaan oikeudemuakisuuden käsitteen muuttumista brittiläisessä julkisessa keskustelussa, sitä voidaan periaatteessa tutkia esimerkiksi lomaalla kansainvälisillä aineistoja (esim. Early Modern English Books Online (EEOB) ja Eighteenth-Century Collections Online (ECCO), jotka yhdessä käsittävät käytännössä kaikki varhaisella uudella ajalla, 1470-1800, Britanniassa painetut kirjat. Hyvä esimerkki muista vastaavista aineistoista on ranskalainen kansallinen Gallica-projekti. Lisäksi erityisesti antiikin aatehistorialliseen tutkimukseen on tarjolla kasvava määrä.^[21] Samalla tähän tekstilouhintaan voidaan yhdistellä luonnolla tavalla käsitteistörin metodeja. Käsitteen muutos voidaan läten ryhtyä tutkimaan aivan eri tavoin kuin käytössä on automatisoituja, laajojen tietoaisteistojen käsitteilyn skaalautuvia mallintamismenetelmiä. Kotimaisena esimerkkinä on kaikkien historiallisten suomalaisten sanomalehtien tekstilouhinta, jossa aineistoista haetaan automatisoidusti pakkamaisia ja heikkeitä.^[22] OCR-menetelmä jonka avulla kuvattu aineisto käännetään koneellisesti suomeksi ei ole täydellinen; parhaimmassakin tapauksessa käänännstarkkuus on noin 90%.^[24] Edellä esitellyn etäluennan kannalta potentiaaliset hyödyt ovat suurempia kuin mahdolliset puutteet, jos olemme kiinnostuneita yleisistä säännönmukaisuuksista ja muutoksista käsitteiden käytössä. Tekstilouhinnan sovellukset ovat kuitenkin vasta muotoutumassa, joten menetelmien tuleva kehitys on vielä vaikea arvioida. Varmaa on kuitenkin, että uusien menetelmien käyttö yleistyy ja muokkaa alan tutkimuskäytäntöjä.

AATEHISTORIAN KANNALLA A RELEVANT TIEN TUTKIMUSAINESTOJEN SAATAVUUDEN ONGELMAT

Tutkimusaineistojen avoimen saatavuuden puutteet muodostavat tutkimuksessa merkittävän pullonkaulan. Tähän liittyy humanistis-yhteiskuntatieteellisessä tutkimuksessa erityisesti kolme keskeistä ongelmaa. Ensinnäkin digitoidut lähdeaineistot ovat usein saatavilla vain kaupallisten lisenssien tai luottamuskäyttöisten yhteistyösopimusten kautta. Toiseksi monet aineistot ovat saatavilla ainoastaan rajatun verkko liittymän kautta, eikä tutkija välttämättä pääse käsiksi koko aineistoon ja alkuperäisiin raakadatoihin joita tarvittaisiin kattavan kokonaisuvan muodostamiseen tutkimusaineistosta. Nykytilanteessa tutkija voi esimerkiksi käyttää edellämainittua, kaikki 1700-luvun painetut brittiläiset kirjat sisältävää ECCO-tietokantaa valmiin käyttöliittymän kautta, mutta kaalis lisenssi tarjoaa vain helpotettuja välineitä yksittäisten dokumenttien sanahakuun tai yksinkertaisiin tilastollisiin työmenetelmiin, ei pääsyä alkuperäiseen raakatekstiin. Tästä on hyötyä lähinnä käyttäjille joilla ei ole varsinaisia tekstilouhinnan tarpeita. Vaikka käyttöliittymien työkaluvalikoima monipuolistui – esimerkiksi GALE-niminen yritys, joka on digitoinut merkittävän määrän historiallisia kirjoja, on kehitellyt 1800-luvun kokoimensa (Nineteenth-Century Collections Online, NCCO) selaamiseen suunnattuja tilastollisia välineitä – tämä merkitsee viime kädessä perinteisten tapojen vilailmistä, joka ei uudista merkittävästi aineistojen kokonaiskäyttöä ja alan tutkimuskulttuuria. Rajoitettujen hakukoneiden käyttö ei konkreettisen lukemisen kannalta eroa yksittäisten kirjojen manuaalisesta selaamisesta, vaikka selaimen yhteistyöjen avulla onkin helpottunut merkittävästi. Tulevaisuuteen katsottuilla tutkijoilla on kuitenkin nopeasti kasvava tarve monipuolisempaan tekstin analyysiin. Yksityiskohtaisempi ja alaa uudistava tekstilouhinta vaatii pääsyä koko alkuperäisaineistoon ja mahdollisuutta sen vapaaseen käsitteilyyn ja jatkojälöstämiseen. Tähän liittyy kolmas aineistojen saatavuuden ongelma, aineistojen käsitteilyn täydistyksen joustavien tieteellisten laskehtikirjastojen puute. Valmiit graafiset käyttöliittymät eivät voi koskaan tarjota samaa joustavuutta aineistojen analyysissä kuin tutkijan itse kirjoittama ohjelmakoodi, jonka laadinnassa voidaan hyödyntää laskehtikirjastojen kautta saatavia valmiita rakennuspalikoita.

Edistystä aineistojen avoimuuden ja laskehtimismenetelmien suhteen on kuitenkin tapahtumassa. Aikaisemmin mainitusta EEOB-kokoelmasta on tehty eri yliopistojen yhteistyössä EEOB-TCF (text creation partnership), jossa OCR-menetelmällä käännettyä tekstiä on korjattu, editoitu ja annotoitu eri tavoin. Vuoden 2015 alusta lähtien EEOB-TCF-tietokanta avottiin yleiseen käyttöön.^[25] Samalla myös tutkijajärjestö esimerkiksi avoimen lähdekoodin ROpenGov-hankkeessa kehitetään tieteellisiä laskehtikirjastoja nimenomaan humanistis-yhteiskuntatieteellisen tutkimuksen tarpeisiin.^[26] Myös valmiita analyysilähteistöja tarjotaan yhä useammin avoimen lähdekoodin toteutuksina. Nähtäväksi jää mikä näiden avustien merkitys tulee olemaan alan kehitykselle. Suljetut EEOB- ja ECCO-tietokannat ovat olleet Suomessa kaupallisesti rajoitetun käyttöliittymän kautta saatavilla jo melkein 15 vuotta, mutta tämän merkitys on jäänyt aatehistorian metodikkaan kannalta vähäisemmäksi kuin mitä voisi toivoa. Samalla, totta kai, alkuperäislähteiden saatavuus on helpottunut moninkertaisesti, mutta aineiston jatkojalostaminen varsinaisia kvantitatiivista tutkimusta kohden ei ole useiden aineistojen kohdalla vielä edes alkanut. Yksi perusteltu väite onkin, että digitaalisten ihmistieteiden ensimmäinen aalto ei ole vielä ehtinyt rantaan asti ennen kuin innoikkaimet ovat alkaneet puhua uudesta. EEOB-TCF:n avaaminen on merkittävä askel, mutta samalla alalla tarvittaisiin laajempaa pääsyä kokonaiseen alkuperäisaineistoihin ja varsinaiseen raakadataan (lähes) tapauksessa kehitettyä tutkimuslaitea. Esimerkiksi suomalaisten sanomalehtien avaaminen yleiseen käyttöön on aatehistoriankin kannalta merkittävää.^[27] Toivottavasti yhä useammat aineistolajit tällaisista instituutionaalisista avauksista rokaistuneina auttavat pilotoimaan uudenlaisia tutkimustapoja.

Huoltella se ei ditoinnilla varustetuilla tietokannoilla on merkittävä lisäarvo, mutta lisäksi olisi tärkeää tutkia edistämättömien alkuperäisaineistojen käytön mahdollisuuksia tutkimuksessa. Nykykäiset tietojenkäsittelymenetelmät tarjoavat mielekkäitä vaihtoehtoja alkuperäisaineistojen automatisoitujen siistämiseen, rikastamiseen ja jalostamiseen samalla kun huomattomaa, mutta tutkimuksen kannalta käyttöhelpoista ja joustavaa dataa on olemassa huomattavasti suurempia määriä kuin huolella viimeistellyä aineistokokelmia. Käytännössä kokemuksemme mukaan on kuitenkin usein erittäin haastavaa saada pääsyä aatehistorian tutkimuksessa keskeisiin alkuperäisaineistoihin edes luottamuskäyttöisen tutkimusyhteistyön kautta, puhumattakaan samojen aineistojen avoimesta julkaisemisesta verkossa.^[28] Keinoitekoisten kaupallisten ja teknisten rajoitteiden yleisyys digitoitujen aineistojen saatavuudessa poikkeaa merkittävästi luonnollisten tutkimuksessa nykyisin yleisistä käytännöistä, jossa mittausaineistojen raakadatat talleennoetaan julkaisutyössä avoimmin tietokantoihin.^[29] Sellitävä nähtäjä ovat erot tutkimuskulttuurissa sekä aineistojen jakelun järjestämisessä ja rahoitusmalleissa.

Luonnontieteellinen tutkimus on ollut edelläkävijä aineistojen avoimessa jakelussa, ja alalla on tiedostettu alkuperäisaineistojen sujuvan saatavuuden merkitys tutkimuksen läpinäkyvyydelle ja avoimuudelle joka on yksi tieteellisen toiminnan kulmakivistä. Aineistojen jakeluun esimerkiksi huoksausyksiössä ja molekyylibiologiassa on perustettu keskitettyjä julkisia tietokantoja, joiden pitkin tähtäimien rahoitus on turvattu.^[30] Humanistis-yhteiskuntatieteellisen tutkimuksen lähdeaineistot on usein kotoi kauppalaisten mallien perusteilla tai esimerkiksi kirjastojen toimesta siirtä ajatuksella, että niiden käyttöoikeuksia voidaan myydä. Tämä johtuu osittain siitä, ettei avoimen julkaisemisen ole perinteisesti nähty tuovan merkittävää lisäarvoa. Tulevaisuudessa tekstimasoinsta voitaisiin kuitenkin esimerkiksi tunnistaa ohjaamattomasti erilaista käsitteitä ilman ennalta asetettuja hierarkkisia kategorioita, jotka ovat aina tavalla tai toisella hallittavissa. Näin perustavanlaatuiset aatehistorian metodologiset kysymykset, kuten ajattelun ja ajattelun kontekstuaalisuutta ja kosellikkailaisten käsitteistörin voivat tulla vna vuosikymmeninä nähtävyytä uudenlaisessa perspektiivissä. Tämä vaatii, että substanssi- ja menetelmäosaaminen löytävi toisensa ja aatehistorian tutkimuskysymysten ehtoista ole enää tutkimus voi ammentaa kieliteknologioista, tilastollisista koneoppimisesta ja muista informaatiotieteiden traditiista. Samalla olisi äärimmäisen tärkeää, että yhteistyötä tutkimuksen ja raakadatan useimmissa tapauksissa hallinnoivien kirjastojen sekä muiden muistiorganisaatioiden kesken syvennetään huomattavassa määrin.

Tieteellisen laskennan ja aineistojen jakelun suunnattujen teknologioiden ja palveluiden kehittämisen myötä tutkimusaineistojen suljetut ja kaupalliset jakelumatit ovat nopeasti vanhentumassa samalla kuin uudet menetelmät tarjoavat tehokkaita mahdollisuuksia digitaalisten aineistojen raakadatan hyödyntämiseen tutkimuksessa. Tutkimuksen avoimuudelle esitetyt vaatimukset ovat myös vahvistuneet. Niin tutkimuksen rahoittajat, julkaisijat, suuri yleisö, kuten myös valvettuneet tutkijat vaativat yhä useammin koko tutkimusprosessin avaamista tavalla joka sisältää tutkimuksen yleisen kuvaamisen lisäksi myös pääsyn alkuperäiseen raakadataan ja analyysikodeihin.^[31] Myös digitaalisten aineistojen jakeluun on tarjolla tehokkaita avoimen lähdekoodin tietokantoja ja rajapintarakaisuja tutkimuslaitosten tarpeisiin. Yksittäiset tutkijat voivat julkaisista dataa hajautetuissa versiohallinta- ja jakelupalveluissa. Myös lavanomaisia kirjoja tai tutkimusartikkelite kirjoitetaan erityisesti yhteistyöprojekteissa kasvavassa määrin tällaisten palveluiden kautta, joista esim kirjat tarjoavat vaikkapa Authora ja GitBook-palvelut.^[32]

Avoimpi tutkimuskulttuuri onkin vauhdilla leviämässä myös humanistis-yhteiskuntatieteelliselle alalle. Jos tahtoa löytyisi, alan lähdeaineistojen julkaiseminen voisi olla monessa tapauksessa helpompaa kuin luonnontieteen alalla, koska aineistot ovat kootaan huomattavasti pienempiä ja esimerkiksi historiallisten tekstiaisteistojen kohdalla niiden liittyy vähemmän yksityisyyden nautustajia tai tekijöidenkysymyksistä nousevia rajoitteita tai salausvaatimuksia kuin vaikkapa bioläketieteiden aineistoihin. Digitaalisten alkuperäisaineistojen julkaiseminen onnistuisi pienemmällä taloudellisella ja teknisellä panoksella kuin luonnontieteellisten aineistojen kohdalla. Aineistojen tutkimuskäytön tarpeen kaupallinen tai tekninen rahoittaminen onkin ongelma, jonka ratkaisemiseen alalla tulisi kiinnittää erityistä huomiota. Ei riitä, että aineistojen hallinnasta vastaavilla taholla on reitorikka kunnossa ja tutkijat kokevat voitiltaan lauseet jotka kirjoittaa rahoitushakemuksen aineistojen ja tutkimusdatan avoimuutta koskevaan kohtaan. Tarvitaan myös konkreettisia toimia.

Avoim data^[33] viittaa digitaalisten aineistojen osalta siihen, että aineisto on saatavilla kokonaisuudessaan, maksutta tai minimaalisin tuotantokustannuksin kone luetavassa muodossa, eikä sen jatkoikätyölle ole asetettu teknisiä, juridisia, kaupallisia tai muita rajoitteita. Aineisto voi siis jaloata, yhdistellä muihin aineistoihin, jakaa edelleen ja käyttää haittamallaan tavalla. Ainoa ehto avoimen aineiston käytölle on viittaaminen alkuperäiseen lähteeseen tavonomaisten tieteellisen käytännön mukaisesti. Tämä avoimen datan määritelmä on sopusoinnussa tieteen avoimuuden kanssa ja siitä tehtävät poikkeamat heikentävät samalla koko tutkimusprosessin avoimuutta ja arvoa. Aineistojen avoimen saatavuuden puolesta voidaan argumentoida erityisesti silloin kun aineistojen keruu ja ylläpito on toteutettu julkisin varoin, esimerkiksi tutkimuslaitosten tai kirjastojen toimesta. Joillakin muilla alalla tehtyjen arvioiden mukaan avoimuuden tuottama lisäarvo tutkimukselle ja yhteiskunnalle saattaa olla merkittävästi suurempi kuin avoimuuden edistämiseen käytettävä taloudellinen panos.^[34] Tämä on suora vastaus kysymykseen kuka maksaa. Avoimen datan hyödyt tieteelle ja yhteiskunnalle ovat ilmeisiä – eräassa äskettäin julkaisussa katsausartikkelissa pyydettiin osuvasti pohtimaan vaikutuksia molekyylibiologian kehitykselle, mikäli ihmisen geenikartasta olisi päätetty julkaista pelkästään kiinnostavat osat.^[35] Sama pätee digitaalisten ihmistieteiden alalla – tällaisten perustavan laatuisten kysymysten pohdinta kuuluu koko tutkijayhteisölle – ei yksittäiselle aineiston tarjoajalle. Tutkimuksen kannalta arvokkaita aineistoja on toki myös kaupallisesti saatavilla. Esimerkiksi Google Books tarjoaa merkittäviä tekstiresursseja, mutta sen sisältämä aineisto ei ole avoimesti saatavilla. Kaupallisten toimijoiden osalta avoimuuden vaatimien on kuitenkin ongelmallisempaa kuin julkisten toimijoiden kohdalla. Tutkimuksen kannalta keskeisten digitaalisten aineistojen avaaminen tutkimuskäyttöön alkuperäisten raakatekstin ja täydellisten kuvallistietojen osalta on jolla tapauksessa tärkeää humanistis-yhteiskuntatieteellisen alan uusien tutkimusmahdollisuuksien edistämiseksi.

AVOIMEN TIET EEN JA T UTKIMUSPROSESSIN LÄPINÄKYVYYDEN MERKIT YS AATEHISTORIAN TUTKIMUKSELLE

Avoimen datan lisäksi keskeistä tutkimuksen läpinäkyvyydelle, toistettavuudelle ja edelleenkäytölle on lisäksi tilastollisten analyysien taustalla olevien lähdekoodien avoin julkaiseminen.^[36] Avoimesta lähdekoodista keskusteleminen saattaa tuntua aatehistorian kontekstissa vielä tänään vieraalta, mutta erilaisten laskehtimien menetelmien käyttö lisääntyessä tulee olemaan merkittävä, jos tutkimuksessa keskitetään alusta lähtien avoimiin toimintatapoihin sen sijaan, että käytettäisiin mahdollisesti valmiita ja rajoitettuja ohjelmistoja. Tämänkäytännön on nopeasti yleistynyt erityisesti luonnontieteen alalla ja vertaisarvioitujen tutkimusten laadintaan liittyyvää lähdekoodia julkaisutaan hajautetuissa versiohallintapalveluissa normaalia käytännönä.^[37] Tämä tapaa parhaimpaan tapauksessa, että jokainen tutkimuksessa tehty tilastollisen analyysin yksityiskohta alkuperäisestä raakadatasta lupailtiin yhteenvetoihin ja visualisointiin on täydellisen läpinäkyvä. Se myös vähentää mahdollisuuksia virheid en piilotteluun ja "tilastokalesteluun", jossa tilastollisen merkitsevyyden kunnollinen testaus jätetään vahingossa tai jopa tahallisesti huomiotta seluotuksessa laajoja tietoaisteistoja.^[38] Tämä vähentää myös tutkimonmaisten työvälineiden toistoa ja potentiaalisesti skaalautuu äärimmäisen hyvin muiden vastaavien aineistojen tutkimukseen. Aatehistoriallisiin lähteisiin liittyen esimerkiksi kirjasto- ja teollisten kuvallistiedot ovat varsin yhtenäisiä ylläpitäjästä riippumatta. Tämä tarjoaa mahdollisuuksia eri maiden kirjasto- ja teollisten yhdistämiseen ja tiedonmuuttamoon tutkimiseen kansainvälisellä tasolla. Avoimen toimintamallin sivutuoteena syntyy perinteiset rajat ylittävät yhteistyötä ja uudenlaisia tutkimuskäytäntöjä.

Tutkimusaineistojen ja menetelmien avoimuus parantaa tehokkuuden lisäksi myös tutkimuksen tiedotavuuatta. Nykytilanteessa tutkija saattaa tehdä suljetun käyttöösi liittyvien avulla monimuukaisiakin tietokantahakua ja muodostaa niiden perusteella erilaisia pääteilytyje tiedojen termien käyttöön historiallisesti kehityksiä, mutta hausta, niiden sisältöä ja pääteilytyje ei välttämättä jää edes tutkijalle itselleen selkeää kirjainpitoa. Analyysin automatisointi tarjoaa ongelmaan ratkaisun. Samalla muutkin tutkijat voivat hyödyntää avointa lähdekoodia omassa työssään ja kehittää sitä halutessaan edelleen omin tarpeisiinsa ja jakaa työnä tulokset takaisin yhteisön käyttöön. Globaali tiedeyhteisön näkökulmasta tällainen malli voi vähentää merkittävästi päällekkäistä työtä ja tarjota mahdollisuuksia verkostoamiseen, perinteiset institutionaaliset ja maantieteelliset rajat ylittävään yhteistyöhön joka hyödyttää kaikkia alan toimijoita ja kihiytää alan tutkimusta. Hajautettu versionhallinta mahdollistaa laajamman tiedeyhteisön osallistamisen välineiden kehittämiseen ja soveltamiseen. Vastaavaa lähestymistapaa on sovellettu muun muassa bioinformaattien tutkimuksessa menetyskyselyä jo pitkään. Muutos on jo käynnistynyt myös humanistis-yhteiskuntatieteellisellä alalla, mutta uusien käytäntöjen vakiintuminen vie aikaa. On silti erittäin tärkeää, että aatehistorian kohdalla laskennallisten menetelmien suhteen alaa suunnatun ja paria tieteellisiä toimintamallia. Ne ovat omaksuttavat voivat saada edellytysaseman ja tehdä merkittävää pioneerityötä mukaten alan kehitysuuntaa.

Viime aikoina esillä ollut avoimen tieteen käsite korostaakin läpinäkyvien ja yhteisöllisten toimintamallien merkitystä. Perinteisesti avoimen tieteen käsitte on katsottu viittaavaan mm. tutkimuksen lähdeaineistojen, tutkimusmenetelmien ja tieteellisen julkaisemisen avoimuuteen. Avoimuus sisältää vapauden tallentaa, muokata ja jakaa edelleen tutkimusmateriaalia tavomaisten tieteellisten viittauskäytäntöjen mukaisesti ilman muita rajoituksia. Tutkimusaineistojen koon kasvassa ja niiden käsitteilyyn tarvittavien informaatiotekniikan menetelmien monimuukaisuudessa tutkimuksen avoimuus on kohdattu uudenlaisia haasteita. Näihin luonnontieteissä vakavasti otettuihin kysymyksiin ei ole kuitenkaan samalla tavalla vielä kiinnitetty huomiota ihmistieteiden tutkimustradition piirissä joka on avoimen käytäntöjen osalta jäänyt kansainvälisesti jälkeen luonnontieteistä. Kun luonnontieteen alalla tutkimuslaitokset jopa kirjaavat avoimen tietokantojen laadulla^[39], humanistis-yhteiskuntatieteellisen tutkimuksen kannalta merkittäviä aineistoja ylläpitävät laitokset vaikuttavat toisinaan välinpitämättömyiltä aineistojen avoimen tutkimuskäytön edistämisen suhteen. Myös kannustimet ovat perinteisesti olleet heikkoja, mutta viime aikoina monet rahottajat ja julkaisijat ovat alkaneet edellyttää uutta julkaisemista tukevaa aineistohallintaa. Tutkimustoiminnassakin voi tulla kysymyksiä kirjain kysymyksiin ja jopa virheellisiin käsityksiin tutkimusaineistojen ja julkaisemisen avoimuuteen liittyen. Tämä on helmasynyt humanistis-yhteiskunnallisessa tutkimuksessa pitäen sisällään aatehistorian. Jos tutkijat eivät itse osaa vaatia aineistojen ja tutkimustulosten vapautta saatavuutta, sitä ei voitane otetta aineistojen ylläpitäjiltäkään. Aineistojen hallinnointi tahot, kunnat, kirjasto- ja muistioorganisaatiot, tulisi sitoa lähdeaineistojen ja menetelmien avaamisen osalta tiiviimmän mukaan tutkimukseen. Lisäksi tulisi tutkia mahdollisuuksia hyödyntää viranomaisen ja muiden tahojen kasvavassa määrin tarjoamien avoimen datan voimavarojen hyödyntämistä tutkimuksen tukena. Argumentti siitä miten tähän parhaiten päästään, on tiivistää merkityksellistä yhteistyötä tutkimuksen ja sitä tukevien organisaatioiden välillä.

Avoimet käytännöt voivat uudistaa tieteenalan toimintakultuuria ja suuntaa. Kansainvälisessä kilpailussa on pystyttävä tekemään työtä yhteistyössä, jossa yhteisten pelisääntöjen kautta pyritään avamaan koko tutkimusprosessi niin että samoista tai samankaltaisista kysymyksistä kiinnostuneiden tutkijoiden ei tarvitse keskiä pyöriä uudestaan, vaan voidaan yhteisvoimin kehittää rikkaampia ja laadukkaita kokonaisuuksia. Esimerkiksi tekstiluonnin osalta kotimaisten tutkimusryhmien kannattaisi tehdä yhteistyötä suomalaisten sanomalehtiaineistojen jaloitamisessa ja analoimisessa. Jos niukat tutkimusresurssit eivät valuisi hukkaan tutkimusryhmien tehossa päällekkäistä työtä. Humanistilla alalla puhutaan usein avoimesta julkaisemisesta, mutta vähemmän tutkimusaineistojen ja tutkimusprosessien avoimuudesta, vaikka nämä saattavat olla tutkimuksen kehittymisen kannalta jopa vielä keskeisempiä. Paradoksaalisesti tätä merkitystä ei aina havaita, kun ajatellaan, että tiedeyhteisön kannalta hyödyllisin julkaisukäytäntö on jollainen, useaan kertaan punloitu ja viilattu tutkimustulos. Aatehistoria on tässä asioiden viimeistelyssä esittänyt ajoittain jopa humanistisessa kontekstissa yllyvää pikkurukkuutta. Kuitenkin myös tutkimusmenetelmät ja tutkimusprosessin aikana käytävä ajatusvaihto ovat keskeisiä tutkimuksessa. Näiden avoin jakaminen tiedeyhteisön kesken tarjoaa laadun tulojen lisäksi myös arvokasta materiaalia ja välineitä joiden pohjalta voidaan tehdä tehokkaasti rakentaa uutta tutkimusta. Avoimia käytäntöjä omaksuneilla tieteenaloilla on hyödynnetty linux-maailmasta tuttua ”julkaise aikaisin ja usein”-periaatetta, jonka mukaan koko kvantitatiivinen tutkimusprosessi ja sen eteneminen avataan yleisölle.^[40] Avoimuuden merkitys tutkimusetiikan ja metodoiden ymmärtämisessä on keskeinen. Näin tutkimusprosessi ja -tulokset tiukkuivat sujuvamminkin tutkijoiden ja tutkimusalojen välillä ja niitä voidaan paremmin hyödyntää innovaatiotoiminnassa, opetuksessa ja kansalaisyhteiskunnassa.

TERVE SKEPTISISMI

Aatehistorian piirissä on oltu pidettyäisiä digitaalisten aineistojen käyttöön ja tiedonluuhintaan koneavusteisin menetelmin samalla kuin kieliteknologia ja siihen liittyvät laskentamenehtimet ovat kehittyneet jo ainakin kolmekymmenen vuoden ajan humanistisen tutkimustradition puitteissa.^[41] Perinteisessä aatehistorian tutkimuksessa yksittäinen sankarintuio lukee ja kirjoittaa, eikä automatisoinnilla tai tieteellisesti laskennalla nähdä juurikaan lisäarvoa. Toisaalta aatehistorian tutkijat eivät ole välttämättä olleet tietoisia uusien informaatiotieteiden menetelmien mahdollisuuksista. Itenäistä laajaajan asiantuntumukseen on pohjautuvaa luovaa ajattelua mikään kone ei tee tutkijan puolesta. Tutkimusta voitaisiin kuitenkin monissa olosuhteissa ottaa käyttöön, joskus tämä saattaa vaihtelevasti olla myös puhdasta sirkusakrobatiasta. Samalla on selvää, että informaatiotieteisiin ja tieteelliseen laskentaan liittyvän menetelmäosaamisen puutteita ihmistieteiden ja erityisesti humanismin tutkimusperinteessä on paikattava. Tämä ei missään muodossa ole helppo tehtävä, koska täysverisistä laskennallisen tieteen ammattilaisista on pulaa kaikilla sovelletuilla. Aatehistoriaan liittyvien tieteellisen laskennan sovellusten viive saattaa selittyä sillä, että aatehistorian ydinosaajat eivät ole läheneet menetelmiä kehittämään ja näin ei ole muodostunut aatehistorian tarpeisiin vastaavia tietojenkäsittelyä hyödyntäviä menetelmiä. Uusi tutkimusala tarjoakin lupaavia mahdollisuuksia ja tuoreita tutkimuskohteita myös menetelmätutkijoille. Haasteena on, että laskennallisten menetelmien edustajat eivät välttämättä osaa ennakoida aatehistorioitsijoiden tarpeita. Sovellettavia menetelmiä löytyy, mutta tutkimuskulttuurit eivät ole vielä kohdanneet merkittävällä tavalla.^[42] Uusien lähestymistapojen ilmetessä tutkimuskenttään korostuu samalla tarve korostaa humanistisen tutkimustradition omien tutkimuskysymysten merkitystä ja ylipäänsä hahmottaa mikä on aatehistoriallinen perustutkimuksen ydin. Mukikasta laskentamenehtimet ja hienot visuaalisoinnit eivät yksinään riitä vastaamaan sovellusten keskeisiin kysymyksiin. Toisaalta vaikka teoreettiset keskustelut ovat tärkeitä, käytännön tulokset ja ymmärrys sanelevat tutkimuksen lopullisen arvon. Mikään tutkimusmenetelmä tai aineisto maailmassa ei korvaa hyviä tutkimuskysymyksiä ja -hypoteeseja. Tutkimuskysymyksen määrittely ja siihen vastaaminen saatavilla olevien aineistojen ja tutkimusmenetelmien perusteella ratkaisee viime kädessä onnistumisen aatehistorian saralla. Jos tutkimusta tehdään liian aineisto- tai menetelmäihäisesti, sen vaikutavuus voi helposti jäädä näennäiseksi.

Humanistisessa ja historian tutkimustraditiiossa monitieteisten ja ennakkoluulottomasti uusia informaatiotieteiden menetelmiä perinteisen kysymysten ratkaisemisessa hyödyntävät tutkijaryhmät ovat yhä harvinaisia.^[43] Korkeittainen ongelma on myös, että useat digitaaliset ihmisteiden hankkeet keskittyvät perinteisiin aiheisiin evätkä tutkimustradition keskuössä olevat henkilöt kykene ymmärtämään digitaalisen komponentin lisäarvoa ennen kuin tutkimuksella on vaikuttavuutta tieteenalan perusongelmien ratkaisemisessa. Lisäksi monet visuaalisoinniseen liittyvät ratkaisut ovat välttämättä tulla varsinaisia tuloksia, vaan keskittyvät tapan muihin otteisiin. Joskus tämä saattaa vaihtelevasti olla myös puhdasta sirkusakrobatiasta. Samalla on selvää, että informaatiotieteisiin ja tieteelliseen laskentaan liittyvän menetelmäosaamisen puutteita ihmistieteiden ja erityisesti humanismin tutkimusperinteessä on paikattava. Tämä ei missään muodossa ole helppo tehtävä, koska täysverisistä laskennallisen tieteen ammattilaisista on pulaa kaikilla sovelletuilla. Aatehistoriaan liittyvien tieteellisen laskennan sovellusten viive saattaa selittyä sillä, että aatehistorian ydinosaajat eivät ole läheneet menetelmiä kehittämään ja näin ei ole muodostunut aatehistorian tarpeisiin vastaavia tietojenkäsittelyä hyödyntäviä menetelmiä. Uusi tutkimusala tarjoakin lupaavia mahdollisuuksia ja tuoreita tutkimuskohteita myös menetelmätutkijoille. Haasteena on, että laskennallisten menetelmien edustajat eivät välttämättä osaa ennakoida aatehistorioitsijoiden tarpeita. Sovellettavia menetelmiä löytyy, mutta tutkimuskulttuurit eivät ole vielä kohdanneet merkittävällä tavalla.^[45] Uusien lähestymistapojen ilmetessä tutkimuskenttään korostuu samalla tarve korostaa humanistisen tutkimustradition omien tutkimuskysymysten merkitystä ja ylipäänsä hahmottaa mikä on aatehistoriallinen perustutkimuksen ydin. Mukikasta laskentamenehtimet ja hienot visuaalisoinnit eivät yksinään riitä vastaamaan sovellusten keskeisiin kysymyksiin. Toisaalta vaikka teoreettiset keskustelut ovat tärkeitä, käytännön tulokset ja ymmärrys sanelevat tutkimuksen lopullisen arvon. Mikään tutkimusmenetelmä tai aineisto maailmassa ei korvaa hyviä tutkimuskysymyksiä ja -hypoteeseja. Tutkimuskysymyksen määrittely ja siihen vastaaminen saatavilla olevien aineistojen ja tutkimusmenetelmien perusteella ratkaisee viime kädessä onnistumisen aatehistorian saralla. Jos tutkimusta tehdään liian aineisto- tai menetelmäihäisesti, sen vaikutavuus voi helposti jäädä näennäiseksi.

LOPUKSI

Kuvailutietojen, tekstinluonnin ja informaatiotieteiden yhdistely tarjoaa uusia tapoja ymmärtää esimerkiksi painotuita fyysisinä objekteina, niiden tuottamiseen liittyviä näkökulmia, sekä laajempia käsitteiden muutoksia.

Digitaalisten aineistojen osalta aatehistorian tutkimuksen kannalta käyttökelpoisia aineistoja on jo hyvin tarjolla. Esimerkiksi Suomen kansalliskirjaston kokoelmat ovat tällaisia. Kotimaiset aineistot tarjoavat kiintoisan ja luontevan tulkinnan aatehistorian tutkimukselle. Lisähaasteena tutkimuksessa on kehittää menetelmiä joilla tekoaineistoita yhdistävä ja erottavia piirteitä voitaisiin sujuvasti analysoida yli kielirajojen. Suomen kaksikielisyys tarjoaa myös erinomaisia mahdollisuuksia pilotoida rajatunmassa mittakaavassa eri kielten samankaltaisista käyttöä ja yhdistelyä tekstiluonnin kanssa. Suomalaisen sanomalehtien tutkimuksessa voi tulla hyödyksi monia kotimaisen aatehistorian osaamisen parhaita puolia liittyen aineiston sisäiseen, käsittehistoriaan ja lingvistiikkaan. Lisäksi tämä tarjoaa erinomaisia mahdollisuuksia kehittää yhteistyötä tieteellisen laskennan asiantuntijoiden kanssa tavalla joka mahdollistaa uusien tutkimusideoiden syntymisen sekä ihmistieteiden tutkimukseen edetä tieteellisen laskennan teorian ja menetelmien kehittämiseen. Ihmistieteiden saralla seuraavaksi tavoitteeksi voitaisiin ottaa pyrkimys kvantitatiivisen tutkimusmenetelmien ymmärtämiseen ja yhteistyön edistämiseen luonnontieteissä jo vakiintuneen avoimen toimintamallin keinoin. Samalla tulee kehittää alan opetusta, jotta pystymme jatkossa auttamaan uusia sukupolvia haastamaan skinnereitä ja kokeilemaan perustellia pyrkä optimoimaan päätyömlään, rukkaveloiaan ja liikuntatutkimuksiaan. ”Omia data”-slikkeä edisteltään henkilökohtaisen ja toisinaan luottamuksellistekni keisteriteitojen saatavuutta omaan henkilökohtaiseen käyttöön. [Takaissin]

Esim. <http://orgprints.org/handle/document/2015/01/31/onko-laske-nnallinen-yhteiskuntatiede-viimein-saapumassa-suomeen/> [Takaissin]

David Lazer et al. ’Computational Social Science.’ Science 323(5915), pp. 721-723, 2009 [Takaissin]

Hans-Georg Gadamer, *Reason in the Age of Science*, trans. Frederick G. Lawrence. Cambridge, MA: MIT Press, 1981. [Takaissin]

Lähtökohdaisesti esim. digitaalisia ihmistieteitä edustavan Dariahin (<https://www.dariah.eu/about/mission.html>) ja kieliteknologian perinteestä kasvavan Clarin-



1. Kiitämme Jani Marjasta, Niko Ilomäkeä, Kaius Sinnemäkeä, Sirkka Havua, Eeva Murtomaata sekä kahta referee-lausunnon kirjoittajaa. Tekstin runko perustuu vuoden 2014 lopulla Oulussa pidettyyn esitelämään. Kiitokset Jouni-Matti Kuukkaselle sekä Oulun yliopiston historian filosofian tutkimuskeskuksen yleisölle stimuloivasta keskustelusta digitaalisten aineistojen mahdollisuuksista. [Takaissin]
2. Kielen ja kuvan yhdistämisestä suomalaisessa tutkimuksessa, ks. Tuomo Hiipälän väitöskirja, Modelling the structure of a multimodal artefact (2013): <https://helda.helsinki.fi/handle/10138/4736>. [Takaissin]
3. Laajempi katsaus aiheeseen löytyy tämän lehden sivuilta: <http://www.ennenjany.net/2015/02/koodaamisen-j-kirjoittamisen-vuoropuhelu-mita-on-digitaalinen-humanistinen-tutkimus/> [Takaissin]
4. Esimerkki uuden profiilin tutkimuskeskuksesta on Sussex Humanities Lab, jossa painopiste on muualle kuin perinteisessä tekstin ja kielen tutkimuksessa. <https://humslab.wordpress.com> [Takaissin]
5. Ihmistieteiden alaa sivuava ajankohtaisia trendejä ovat esimerkiksi ”biohakkerointi”, jossa henkilö monitoroi elintoimintojaan kirjainpöydillä ja mittailaittein ja voi saada tiedon perustellia pyrkä optimoimaan päätyömlään, rukkaveloiaan ja liikuntatutkimuksiaan. ”Omia data”-slikkeä edisteltään henkilökohtaisen ja toisinaan luottamuksellistekni keisteriteitojen saatavuutta omaan henkilökohtaiseen käyttöön. [Takaissin]
6. Esim. <http://orgprints.org/handle/document/2015/01/31/onko-laske-nnallinen-yhteiskuntatiede-viimein-saapumassa-suomeen/> [Takaissin]
7. David Lazer et al. ’Computational Social Science.’ Science 323(5915), pp. 721-723, 2009 [Takaissin]
8. Hans-Georg Gadamer, *Reason in the Age of Science*, trans. Frederick G. Lawrence. Cambridge, MA: MIT Press, 1981. [Takaissin]
9. Lähtökohdaisesti esim. digitaalisia ihmistieteitä edustavan Dariahin (<https://www.dariah.eu/about/mission.html>) ja kieliteknologian perinteestä kasvavan Clarin-

- verkoston (<http://www.clarin.eu/content/mission>) välillä nähtiin jännitteitä, mutta kansainvälinen suuntaus on tällä hetkellä näiden verkostojen luonnollinen yhdyntyminen. [Takaisin]
10. ks. myös <http://widerscreen.fi/numero/2013-2-3/kieltäydyn-maarittelema-digitaalista-kulttuuria/> [Takaisin]
11. Franco Moretti, 'Conjectures of world literature', *New Left Review*, 1, 2000, 54-67 and Moretti, 'More conjectures', *New Left Review*, 20, 2003, pp. 73-81. Morettin tekstejä on kerätty kokoelmaan: *Distant Reading*, Verso, 2013. [Takaisin]
12. Erinomaista osoituksena laajojen aineistojen etäluennan käyttömahdollisuuksista: Sara Klingenstein, Hitchcock, Tim and DeDeo, Simon, 'The civilizing process in London's Old Bailey', *Proceedings of the National Academy of Sciences*, 111, 2014, pp. 9419-9242. Ks. myös <http://www.oldbaileyonline.org/> [Takaisin]
13. Esimerkki tällaisesta menestyksekkäästä korpus-tutkimuksesta on Helsingin yliopiston englanninkielen variaation tutkimusyksikkö, VARiENG, <http://www.helsinki.fi/vareng/> [Takaisin]
14. Quentin Skinner, *Visions of Politics Volume 1. Regarding Method*, Cambridge University Press, 2002. [Takaisin]
15. Skinner, *Visions of Politics*, s. 57-102, 169. [Takaisin]
16. Reinhart Koselleck, 'Einführung', Otto Bruner, Werner Conze & Reinhart Koselleck (hrsg.), *Geschichtliche Grundbegriffe. Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*. Band 1, Stuttgart, Klett-Cotta 1972, s. XIII-XXVII. Reinhart Koselleck, *Vergangene Zukunft. Zur Semantik geschichtlicher Zeiten*, Frankfurt am Main, Suhrkamp 1979. Kari Palonen on, aivan oikein, alle viivannut skenaarilaisen kontekstuaalismin yhteyttä käsitteistoriaan, Palonen, 'The History of Concepts as a Style of Political Theorizing: Quentin Skinner's and Reinhart Koselleck's Subversion of Normative Political Theory', *European Journal of Political Theory*, 1, 2002, pp. 91-106. Katso myös, Melvin Richter, 'Toward a Lexicon of European Political and Legal Concepts: A Comparison of Begriffsgeschichte and the "Cambridge School"', *Critical Review of International Social and Political Philosophy*, 6, 2003, pp. 91-120. Esimerkkinä eteenpäin katsovalta käsitteistoriaprojektilta voidaan mainita Bernhard Jusseinin johtama keskiäkahanke, [https://github.com/OnOffennica](http://www.geschichte.uni-frankfurt.de/43010469/jussenb; Mehler, Alexander; Schwandt, Silke; Gleim, Rüdiger; Ernst, Alexandra. Inducing linguistic networks from historical corpora. Towards a new method in historical semantics, in: Paul Bennett, Martin Durrell, Silke Scheible, Richard J. Whitt (Hrsg.), New Methods in Historical Corpora, Tübingen 2013, S. 257-274. Kiitämme tässä erityisesti Jani Marjasta käsitteistoriaprojektin nykytilaan liittyvästä katsauksesta. [Takaisin]</p><p>17. Jouni-Matti Kuukkanen on tosin ehtinyt jo ryhtyä tähän jo ennen kuin on alettu pohlia ajatuksia uusien koneavusteisten menetelmien mahdollisuuksista: Jouni-Matti Kuukkanen, 'Making Sense of Conceptual Change', <i>History and Theory</i>, 47, 2008, pp. 351-372. [Takaisin]</p><p>18. <a href=). Tässä julkaisussa Fennicaan liittyvät graafit ovat Niko Ilomäen laatimia. [Takaisin]
19. Mielenkiintoista kehitystä ja seurattavaa toki riittää, niistä tässä mainittakoon: <http://benschmidt.org/projects/digital-humanities-research/> [Takaisin]
20. Suomalaisen sanomalehtien tekstiutinimassa, suuntaa näyttää tällä hetkellä Tampereen yliopiston Informaatiotieteiden yksikön piirissä tehtävää työ. Myös muita aloitteita mm. Helsingissä ja Turussa on vireillä. [Takaisin]
21. Tämän dynamiikan tutkimuksen idean "isä" on Jani Marjanen tässä meidän välisen keskustelun kontekstissa. [Takaisin]
22. Latinkielisistä aineistosta yleisemmin ks. www.brepolis.net. [Takaisin]
23. <http://digi.kansalliskirjasto.fi/sanomalehti> [Takaisin]
24. Kimmo Kettunen et al. 'Analyzing and Improving the Quality of a Historical News Collection using Language Technology and Statistical Machine Learning Methods.' In: *Proc. of IFLA 2014*, Lyon (2014). [Takaisin]
25. <http://www.textcreationpartnership.org/2014/12/24/eebo-tcp-phase-i-public-release-what-to-expect-on-january-1/> [Takaisin]
26. [ropengov.github.io](https://github.com/ropengov) [Takaisin]
27. FINCLARIN on avannut tätä aineistoa toukokuun 2015 lopulla ladattavassa muodossa vuoteen 1875 asti, <https://korp.csc.fi/download/Digilib-pub/> Kansalliskirjasto tulee todennäköisesti jatkamaan tätä vuoteen 1910 asti lähitulevaisuudessa, jakelekanava saatata hyvin olla kyseinen FINCLARIN ja CSC. [Takaisin]
28. Esimerkiksi Dewey-järjestelmän saaminen tutkimuskäyttöön on osoittautunut erittäin hankalaksi. Samanlaisia kokemuksia liittyy niin ECCO-aineiston raakadatan saamiseen kuin myös eri kirjastoalueiden marc-tietokantojen saamisesta tutkimuskäyttöön. [Takaisin]
29. FS Collins et al. The human genome project: lessons from large-scale biology. *Science* 300(5617):286-290, 2003. [Takaisin]
30. Ks. esim. ihmisen perimän kartoitushanke GenBank <http://www.ncbi.nlm.nih.gov/genbank/> ja Euroopan hiukkasfysiikan tutkimuskeskuksen CERNin avoimen datan portaalit: <http://opendata.cern.ch/> [Takaisin]
31. <http://www.sciencemag.org/content/336/6078/159.summary> [Takaisin]
32. <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001747> [Takaisin]
33. <https://www.authorea.com/>; <https://www.gitbook.com/> [Takaisin]
34. <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195> [Takaisin]
35. <http://www.worldbank.org/content/dam/Worldbankdocument/Open-Data-for-Economic-Growth.pdf>; http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation [Takaisin]
36. Mark Watson. When will 'open science' become simply 'science'? *Genome Biology* 16:101, 2015. [Takaisin]
37. <http://www.sciencemag.org/content/336/6078/159.summary>; <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001747> [Takaisin]
38. Esim. Github ja Bitbucket [Takaisin]
39. <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002106> [Takaisin]
40. <http://bib.oxfordjournals.org/content/2/2/143.short> [Takaisin]
41. Eric S. Raymond (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media. ISBN 1-56592-724-9. [Takaisin]
42. Helsingin yliopistossa esim. Kimmo Koskenniemi. [Takaisin]
43. Tästä ks. <http://www.geschichte.uni-frankfurt.de/43010469/jussenb> ja http://historymanifesto.cambridge.org_, sekä David Armitage, 'What's the Big Idea? Intellectual History and the Longue Durée', *History of European Ideas*, 38, 2012, s. 493-507. [Takaisin]
44. Ks. esim. Pasi Ihalaisen tuotanto ja Jani Marjanen, *Den ekonomiska pattoisomens uppgång och fall: Finska hushållningsläskaper i europeisk, svensk och finsk kontext 1720-1940*, 2013. [Takaisin]
45. Keskiäikaisten käsikirjoitusten tutkimuksessa historian alalla pioneereina Suomessa ovat kuitenkin olleet Teemu Roos ja Tuomas Heikkilä, 'Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets', *Literary and Linguistic Computing*, 24, 2009, s. 417-433. [Takaisin]